

## **II Conferencia Internacional sobre Brecha Digital e Inclusión Social (Leganés, Madrid, del 28-30 de octubre de 2009).**

### **utilización de los sistemas de diálogo hablado para el acceso a la información en diferentes dominios**

**David Griol**

*Universidad Carlos III de Madrid, Dpto. de Informática, Escuela Politécnica Superior,  
Avda. de la Universidad 30, 28911- Leganés (España)*

[dgriol@inf.uc3m.es](mailto:dgriol@inf.uc3m.es)

**Zoraida Callejas**

*Universidad de Granada, Dpto. de Lenguajes y Sistemas Informáticos, Escuela Técnica Superior de Ingeniería Informática y de  
Telecomunicaciones, C/ Periodista Daniel Saucedo Aranda s/n, 18071- Granada (España)*

[Zoraida@ugr.es](mailto:Zoraida@ugr.es)

**Ramón López-Cózar**

*Universidad de Granada, Dpto. de Lenguajes y Sistemas Informáticos, Escuela Técnica Superior de Ingeniería Informática y de  
Telecomunicaciones, C/ Periodista Daniel Saucedo Aranda s/n, 18071- Granada (España)*

[rlopezc@ugr.es](mailto:rlopezc@ugr.es)

**Ana Gutiérrez**

*Universidad de Granada, Dpto. de Lenguajes y Sistemas Informáticos, Escuela Técnica Superior de Ingeniería Informática y de  
Telecomunicaciones, C/ Periodista Daniel Saucedo Aranda s/n, 18071- Granada (España)*

[agcast@correo.ugr.es](mailto:agcast@correo.ugr.es)

**RESUMEN:** La acción de conversar es el modo más natural para resolver un gran número de acciones cotidianas entre los seres humanos. Por este motivo, un interés histórico dentro del campo de las Tecnologías del Habla ha sido utilizar estas tecnologías en aplicaciones reales, especialmente en aplicaciones que permitan a una persona utilizar su voz para obtener información mediante la interacción directa con una máquina o para controlar un determinado sistema. El objetivo es disponer de sistemas que faciliten la comunicación persona-máquina del modo más natural posible, es decir, a través de la conversación. En esta comunicación se resumen los resultados de la aplicación de estas tecnologías para el desarrollo de diferentes sistemas de diálogo en los que la interacción entre el usuario y el sistema se lleva a cabo mediante habla espontánea en castellano. Para su implementación se ha primado la utilización de diferentes herramientas de software libre para el reconocimiento automático del habla, comprensión del lenguaje natural, gestión del diálogo y síntesis de texto a voz. De este modo, el objetivo principal de la comunicación es presentar las principales ventajas que proporcionan los sistemas de diálogo para facilitar el acceso a diferentes servicios dentro de dominios semánticos restringidos, qué posibilidades brinda el uso de herramientas de software libre para su implementación y su evaluación en diferentes casos concretos de aplicación.

**PALABRAS CLAVES:** Sistemas de Diálogo, Procesamiento Lenguaje Natural, Accesibilidad.

#### **1. Introducción**

El diccionario de la Real Academia Española de la Lengua define el diálogo como la plática entre dos o más personas que alternativamente muestran sus ideas o afectos. La acción de conversar es el modo más natural para resolver un gran número de acciones cotidianas entre los seres humanos: obtener una determinada información, contratar un servicio, solicitar un pedido, conocer el estado de un determinado proceso, etc.

Por este motivo, un interés histórico dentro del campo de las Tecnologías del Habla ha sido utilizar estas tecnologías en aplicaciones reales, especialmente en aplicaciones que permitan a una persona utilizar su voz para obtener información mediante la interacción directa con una máquina o para controlar un determinado

sistema. El objetivo es disponer de sistemas que faciliten la comunicación persona-máquina del modo más natural posible, es decir, a través de la conversación.

Un sistema de diálogo puede, de esta forma, entenderse como un sistema automático capaz de emular a un ser humano en un diálogo con otra persona, con el objetivo de que el sistema cumpla con una cierta tarea (normalmente suministrar una cierta información o llevar a cabo una determinada tarea).

Descartando el más simple de los casos, este tipo de aplicaciones requieren una secuencia de interacciones entre la persona y la máquina para conseguir que el usuario consiga su propósito. Así, el objetivo del usuario se alcanza gradualmente con la interacción con la máquina a lo largo de varios turnos de diálogo. Este hecho obliga a dotar al sistema de la complejidad necesaria para que pueda: referenciar durante el diálogo la información que haya aparecido anteriormente, tomar la iniciativa para reconducir el diálogo dentro del dominio en el que se ha definido, solicitar información necesaria para cumplir el objetivo solicitado, requerir aclaraciones cuando exista la duda sobre la información aportada por el usuario, etc.

## 2. Metodología

Construir una aplicación informática que pueda mantener una conversación con una persona de manera natural sigue siendo hoy en día un reto, dada la gran cantidad de fuentes de conocimiento que son necesarias y las limitaciones de las tecnologías utilizadas para obtener información del usuario. No obstante, los constantes avances de la investigación en Tecnologías del Habla han permitido que sean factibles actualmente sistemas de comunicación persona-máquina mediante la voz, capaces de interactuar con cierto grado de flexibilidad (iniciativa mixta en el desarrollo del diálogo). Estos sistemas están siempre orientados a tareas de información muy específica (dominios semánticos restringidos).

Disponiendo de un interfaz hablado, se libera al usuario de utilizar otros canales tradicionales como el teclado, el ratón o la pantalla. Dado este interfaz vocal, el número de entornos y tareas en los que pueden aplicarse estos sistemas es enorme: sistemas que proporcionen información sobre horarios y precios de transportes públicos, sistemas de información meteorológica, servicios de banca electrónica, aplicaciones accesibles desde los vehículos, sistemas que faciliten el acceso a la información a personas con discapacidades, acceso a servicios y control de máquinas vía telefónica (fija o móvil), portales de voz, etc.

La Figura 1 resume las acciones básicas que debe realizar un sistema de diálogo para cumplir la finalidad global para la que fue diseñado.

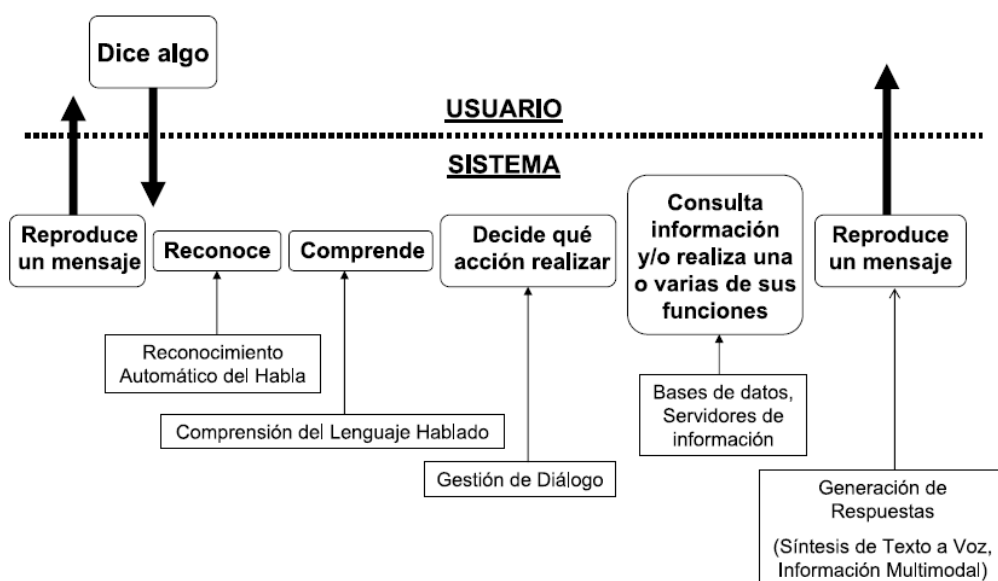


Figura 1. Diagrama de acciones de un sistema de diálogo

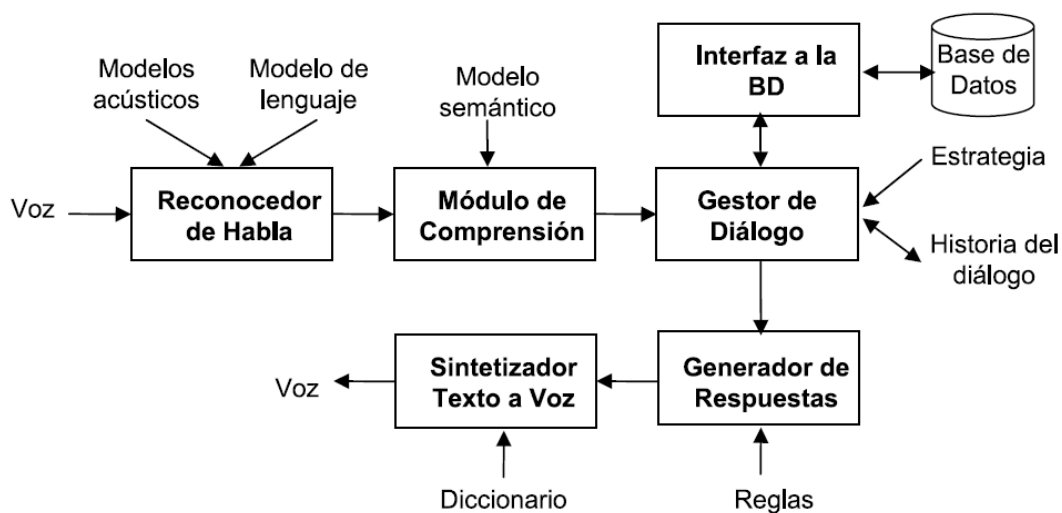
Tal y como se observa en esta figura, el sistema genera un mensaje inicial, normalmente para dar la bienvenida o informar al usuario sobre las características y funcionalidades del sistema. Tras cada intervención del usuario, el sistema debe realizar un conjunto de acciones básicas que se repiten cíclicamente como respuesta a cada acción del usuario:

- Reconocer la secuencia de palabras mencionadas por el usuario.
- Extraer el significado de dichas palabras, es decir, comprender qué información es útil en el dominio del sistema.
- Realizar operaciones de acceso a base de datos u otros recursos del sistema, en los que se almacena la información que solicita el usuario o se registran las operaciones que desea conocer.
- Decidir qué acción o acciones deben realizarse a continuación de cada solicitud del usuario, es decir, qué respuesta debe suministrar el sistema.
- Reproducir un mensaje hablado que informe al usuario qué acción ha seleccionado el sistema.

De este modo, dado el gran número de operaciones que deben realizarse, es habitual un desarrollo modular de los sistemas de diálogo hablado, lo que permite desglosar las dificultades entre los diferentes componentes del sistema. Un sistema de estas características se puede describir en términos de los siguientes módulos:

- **Módulo de Reconocimiento Automático del Habla**, reconoce la señal vocal pronunciada por el usuario y proporciona la secuencia de palabras reconocida más probable (o las k más probables).
- **Módulo de Comprensión del Habla**, a partir de las(s) secuencia(s) de palabra(s) reconocida(s), el sistema obtiene una representación semántica de su significado.
- **Gestor de Diálogo**, considera la interpretación semántica de la petición del usuario, la historia del proceso de diálogo, la información de la aplicación disponible en ese punto y el estado del sistema, y determina la siguiente acción que debe tomar el sistema siguiendo la estrategia del diálogo.
- **Módulo de Consulta a la Base de Datos de la Aplicación**, recibe peticiones de consulta a la base de datos por parte del gestor de diálogo, las procesa y devuelve el resultado al gestor.
- **Módulo de Generación de Respuestas**, recibe la respuesta del sistema en forma de cierta representación formal y tiene como función la generación de una frase, gramaticalmente correcta y en un lenguaje lo más cercano posible al lenguaje natural, que transmita el mensaje generado por el gestor de diálogo. La respuesta del sistema proporcionada por el generador de respuestas puede incorporar otras modalidades de información (vídeo, tablas con datos, gestos a reproducir por un avatar...).
- **Sintetizador de Texto a Voz**, componente que recibe la respuesta del sistema como texto en lenguaje natural y genera la correspondiente señal de audio, que será la respuesta que llegará al usuario.

La Figura 2 muestra la arquitectura modular descrita para el desarrollo de sistemas de diálogo hablado. En cuanto a las metodologías utilizadas para el desarrollo de los diversos módulos que componen el sistema de diálogo, cabe destacar la utilización de técnicas basadas en reglas o la aplicación de métodos estadísticos, basados usualmente en el aprendizaje de un modelo a partir de un corpus de diálogos.



**Figura 2.** Arquitectura modular de un sistema de diálogo hablado

### 3. Objeto de estudio: aplicaciones de los sistemas de diálogo hablado

La complejidad de la interacción entre el usuario y el sistema de diálogo puede variar y algunos de los componentes previamente descritos pueden no ser necesarios. Por ejemplo, para un sistema sencillo, no es necesario un análisis semántico. Sin embargo, para el desarrollo de un compañero conversacional deben utilizarse todos los módulos para interpretar la entrada del usuario, tomar decisiones justificadas sobre qué respuesta debe generar el sistema, y finalmente adaptar la respuesta a las necesidades y expectativas de usuario.

Existe un gran número de tareas que pueden desempeñarse mediante el uso de sistemas de diálogo oral. La más extendida es facilitar información sobre una tarea específica. Algunos ejemplos de aplicación son los siguientes:

- DARPA Communicator - Interfaces conversacionales inteligentes para la reserva de billetes de avión (DARPA, 1992, 1994).
- Voyager - Información turística del área de Boston (Glass et al., 1995).
- ARISE - Sistema de información ferroviaria en distintos idiomas europeos (den Os et al., 1999).
- AUGUST - Sistema de diálogo sueco que emplea un agente animado para dar información acerca de Estocolmo (Gustafson et al., 1999).
- Adapt - Sistema de diálogo multimodal para buscar apartamentos en el mercado inmobiliario de Estocolmo (Gustafson et al., 2000).
- Jupiter - Sistema de información meteorológica por teléfono (Zue et al., 2000).
- Mercury - Sistema de reserva de vuelos e información meteorológica (Seneff y Polifroni, 2000).
- CTT-Bank - Sistema de banca operado oralmente por teléfono (Melin et al., 2001).
- SmartKom - Sistema de diálogo multimodal con diversos dominios de aplicación, entre ellos la reserva de entradas de cine (Alexandersson y Becker, 2001).
- Let's Go - Sistema de diálogo oral para obtener información acerca de los autobuses de Pittsburgh, orientado principalmente a extranjeros y ancianos (Raux et al., 2005).
- Amities project - Interacción multilingüe con servicios e información bancaria (Hardy et al., 2006).
- DisCoH - Sistemas de diálogo oral de ayuda en congresos (Andeani et al., 2006).
- HIGGINS - Navegación y guía de ciudades para peatones (Skantze et al., 2006).
- TALK TownInfo - Sistema de diálogo multimodal para escenarios de información turística (Lemon et al., 2006).
- Conquest - Sistema de diálogo oral que aporta información de horarios en congresos (Bohus et al., 2007).

Los sistemas de diálogo se han empleado también para la educación y el entrenamiento, particularmente con el fin de mejorar habilidades fonéticas y lingüísticas de los usuarios:

- LARRI - Sistema de diálogo multimodal que asiste y guía a personal de aviones de combate durante tareas de mantenimiento (Bohus y Rudnicky, 2002).
- ITSPOKE - Sistemas de diálogo oral de tutorización que conversa con los estudiantes para darles información y corregir sus errores de aprendizaje (Litman y Silliman, 2004).
- Radiobot-CFF - Sistema de diálogo oral que participa en conversaciones que ayudan a entrenar soldados en los procedimientos de inicio de misiones de artillería (Roque et al., 2006).
- VOCALIZA - Sistema de diálogo para terapias de voz en el idioma español que ayuda a los terapeutas a entrenar a usuarios con distintas patologías y carencias en su habilidad lingüística (Vaquero et al., 2006).
- LISTEN - Tutor de lectura que muestra historias en una pantalla y escucha a los niños mientras leen en voz alta (Mostow, 2008).

La interacción oral puede ser la única manera de acceder a la información en muchos casos, como cuando la pantalla disponible es demasiado pequeña para mostrar la información (P.ej. en dispositivos de bolsillo) o cuando los ojos del usuario están ocupados en otras tareas (P.ej. en la conducción):

- MUST - Servicios de información multimodal y multilingüe para terminales móviles de tamaño reducido (Boves y Os, 2002).
- VICO - Virtual Intelligence CO-driver. Permite la interacción natural entre las personas y dispositivos digitales y servicios en el automóvil (Mattasoni et al., 2002).
- Athosmail - Sistema multilingüe y adaptativo para la lectura de correos electrónicos utilizando teléfonos móviles (Jokinen et al., 2004).
- CHAT - Asistente conversacional en tareas dentro del automóvil como por ejemplo operar un reproductor de MP3 (Weng et al., 2006).
- DICO - Sistema de diálogo multimodal que permite al conductor de un vehículo controlar dispositivos y acceder a servicios de Internet mediante la voz (Villing y Larsson, 2006).

La interacción oral es también útil para el control de dispositivos y robots, especialmente en entornos inteligentes:

- WITAS - Interfaz para sostener conversaciones multimodales con el robot-helicóptero WITAS (Lemon et al., 2001).
- ODISEA - Odisea es un sistema de diálogo oral que permite la interacción entre usuarios y entornos inteligentes. Los componentes del diálogo se generan automáticamente y permiten la interacción oral entre el entorno y los usuarios (Montoro et al., 2004, 2006).
- SENECA - Interfaz oral para aplicaciones de entretenimiento, navegación y comunicación en entornos móviles (Minker et al., 2004).
- Clarissa - Navegador manejable mediante la voz, permitiendo a los astronautas ser más eficientes con sus manos y ojos y prestar atención completa a la tarea mientras que navegan con el sistema utilizando comandos orales (Rayner et al., 2005).
- MIMUS - Sistema de diálogo multimodal para controlar un hogar inteligente (Pérez et al., 2006).
- STanford AI Robot (STAIR) - Asistente conversacional robótico para la oficina y el hogar (Krsmanovic et al., 2006).
- Cogniron - El robot compañero cognitivo (Menezes et al., 2007).

Finalmente, los agentes virtuales y compañeros constituyen la aplicación más exigente de los sistemas de diálogo:

- Collagen - Asistentes conversacionales y agentes colaborativos (Rich y Sidner, 1998).

- AVATALK - Diálogos naturales e interactivos con seres humanos virtuales (Hubal et al., 2000).
- COMIC - Diseño de cuartos de baño mediante la interacción con un avatar capaz de generar expresiones faciales (Catizone et al., 2003).
- NICE - Personajes históricos y literarios capaces de comunicarse de forma natural y divertida con niños y adolescentes (Corradini et al., 2004).

Beneficiándose de las mejoras incesantes en las áreas del reconocimiento del habla, el procesamiento del lenguaje natural y de la síntesis del habla, las primeras iniciativas de investigación relacionadas con los sistemas de diálogo oral surgen a principios de los años 80 como resultado de dos grandes proyectos con financiación gubernamental: el programa de sistemas de lenguaje hablado DARPA en los Estados Unidos y el programa Esprit SUNDIAL en Europa. El dominio del programa DARPA fue la consulta de información relativa a vuelos (Air Travel Information Services, ATIS). En el proyecto participaron un conjunto de laboratorios de investigación de Estados Unidos. El principal objetivo de la investigación fue el estudio y desarrollo de las tecnologías relativas al reconocimiento del habla y comprensión del lenguaje, bajo el dominio de la reserva de vuelos utilizando el canal telefónico (DARPA, 1992) (ARPA, 1994). Dado que todos los participantes del proyecto utilizaron la misma base de datos, fue posible comparar el funcionamiento de los diferentes prototipos desarrollados, llevándose a cabo un gran esfuerzo para realizar evaluaciones periódicas de los diferentes sistemas. Actualmente, el corpus de diálogos ATIS sigue siendo un recurso utilizado por desarrolladores y evaluadores de sistemas de diálogo.

La tarea ATIS sirvió de marco para el desarrollo de los primeros proyectos, como los llevados a cabo en AT&T, concretamente el proyecto AMICA (Pieraccini et al., 1997), donde se aplicaron diferentes métodos estocásticos en un sistema de diálogo con iniciativa mixta. ATIS también fue el punto de partida de las investigaciones del MIT y de la CMU, dentro de este proyecto se desarrollaron los sistemas CMU ATIS (Ward, 1991) y MIT ATIS (Seneff et al., 1991) (Zue et al., 1993). El proyecto SUNDIAL (Speech Understanding and Dialogue) (Peckham, 1993), financiado por la Comunidad Europea, tuvo como dominio la consulta de horarios de trenes y aviones en inglés, francés, alemán e italiano. El objetivo del proyecto fue construir sistemas de diálogo en tiempo real capaces de mantener una conversación con el usuario siguiendo una estrategia cooperativa. Además de tratar aspectos de reconocimiento y comprensión del habla, un tema de estudio en el que se centró la investigación fue el modelado del diálogo hablado, desarrollándose diferentes aproximaciones para realizar la gestión del diálogo.

La investigación llevada a cabo en SUNDIAL condujo a un gran número de proyectos con financiación europea centrados en el modelado del diálogo, como VerbMobil (Bos et al., 1999), DISC (Bernsen y Dybkjaer, 1997) y ARISE (Den et al., 1999). También cabe mencionar el sistema Philips (Aust et al., 1994) (Aust et al., 1995), sistema comercial bajo el dominio de consultas sobre horarios de trenes desarrollado a partir de la investigación llevada a cabo en SUNDIAL. Dentro del proyecto europeo ARISE se desarrollaron seis sistemas en paralelo: dos prototipos italianos basados en la tecnología desarrollada por el CSELT (Castagneri, Baggia, y Danieli, 1998) (Baggia, Castagneri, y Danieli, 2000), un prototipo francés desarrollado por el LIMSI (Lamel et al., 2000) (Den et al., 1999) y dos prototipos en holandés y uno en francés basados en la tecnología Philips.

Más recientemente, entre los programas de investigación de mayor importancia, cabe destacar DARPA Communicator. En este programa, con financiación gubernamental y centrado en el desarrollo de tecnologías del habla, participaron centros de investigación en Estados Unidos y Europa. El objetivo establecido fue el desarrollo de una nueva generación de sistemas de diálogo, utilizando como entrada del sistema no únicamente la voz sino existiendo también la posibilidad de incorporar otros tipos de información multimodal. Los sistemas desarrollados dentro de este programa soportan interacciones complejas con el usuario, desde el punto de vista que tanto el sistema como el usuario pueden iniciar la conversación, cambiar de tema o interrumpir al otro participante del diálogo. Los dominios de aplicación (en los que se incluye la planificación de viajes) requieren usualmente el acceso a múltiples fuentes de información, representando un avance con respecto a los sistemas desarrollados en los programas ATIS o SUNDIAL.

Los investigadores de la CMU desarrollaron el sistema Carnegie Mellon Communicator ([www.speech.cs.cmu.edu/Communicator](http://www.speech.cs.cmu.edu/Communicator)) (Rudnicky et al., 1999), que permite obtener información de itinerarios complejos que incluyen reservas múltiples de vuelos, hoteles y alquiler de coches. La arquitectura del sistema se basa en la definición de módulos agentes de dominio que se encargan de la gestión de la información más específica. Paralelamente, investigadores de la Universidad de Colorado desarrollaron el sistema CU Communicator (Ward y Pellom, 1999), que aborda la misma tarea. De los diversos proyectos que siguieron, sobresalen los sistemas implementados por el grupo del LIMSI: el sistema PARIS-SITI, dedicado a facilitar información turística sobre la capital francesa, el sistema MASK (Multimodal Multimedia Service Kiosk) (Gauvain et al., 1995) y el sistema sobre información de trenes RAILTEL (Billi y Lamel, 1997).

Durante la historia de los sistemas de diálogo oral, algunos expertos se han atrevido a prever cuáles serían las líneas futuras de investigación en el área. Estos objetivos se han desplazado gradualmente hacia metas cada vez más complejas. Tal y como se ha reflejado en los resultados de investigación comentados previamente, durante los años 90 las tendencias principales se centraron en aumentar la robustez de los diversos módulos del sistema (RAH, PLN, GD, GLN, CTH) (Cole et al., 1995; Kacic, 1999; Zue y Glass, 2000; Mangold, 2001). A partir de 2003 los expertos han propuesto objetivos de más alto nivel, tales como proveer al sistema de razonamiento avanzado, capacidad de resolución de problemas, facultad de adaptación, proactividad, inteligencia afectiva, multimodalidad y multilingüismo (Dale, 2003; Jokinen, 2003; Gao et al., 2005; Haas et al., 2005; Minker et al., 2006b,a).

De este modo, los expertos prevén sistemas de diálogo inteligentes, adaptables, dinámicos, portables y multimodales. Todos estos conceptos no son mutuamente excluyentes, así la inteligencia del sistema está relacionada con su capacidad para adaptarse a nuevas situaciones y ésta a su vez le confiere mayor portabilidad para facilitar su uso en diversos entornos.

#### **4. Resultados: aplicación práctica de los sistemas de diálogo orales**

A continuación se presentan aplicaciones prácticas de los sistemas de diálogo en las que hemos participado los autores de la presente comunicación.

##### **4.1. El sistema de diálogo DIHANA**

El objetivo del proyecto DIHANA (Griol et al., 2008) fue el estudio y desarrollo de un sistema robusto de diálogo modular y distribuido para el acceso a sistemas de información. En concreto, se profundizó en aspectos metodológicos fundamentales en los campos del tratamiento del habla espontánea, modelado del lenguaje, comprensión y diálogo. La tarea del proyecto es proporcionar información en lenguaje natural sobre servicios, horarios y precios de trenes de largo recorrido en español. A continuación se resumen las características básicas de cada uno de los módulos que conforman el sistema de diálogo desarrollado para el proyecto DIHANA, así como el formato de comunicación adoptado para la comunicación entre los diferentes componentes del sistema:

- Módulo de Reconocimiento Automático del Habla (RAH): Se ha desarrollado un reconocedor automático del habla expresamente para el proyecto. Asimismo, se ha realizado la integración del reconocedor del habla Sphinx-II y el entrenamiento de los modelos acústicos con SphinxTrain, ambos desarrollados por la Carnegie Mellon University.
- Módulo de comprensión (CH): Para la definición de la semántica de la tarea se utiliza el concepto de frame: cada intervención del usuario genera uno o más frames que representan el significado de la intervención.
- Gestor de diálogo (GD): la actualidad hay disponibles cuatro gestores de diálogo adaptados a la filosofía de comunicación del sistema. Dos de ellos están basados en reglas, uno de ellos es estocástico y el último se basa en un modelo mixto.

- Módulo de consulta a la base de datos de la aplicación (MCBD): sistema utiliza una base de datos en PostGreSQL.
- Módulo de generación de respuestas: La estrategia que sigue el sistema consiste en la utilización de plantillas de respuestas adaptadas a los tipos de consulta.
- Módulo de síntesis texto-voz: Se ha integrado el sistema de síntesis de voz Festival desarrollado por la Carnegie Mellon University y la University of Edinburgh.

Tal y como puede observarse, uno de las principales objetivos utilizados para el desarrollo del sistema fue la utilización de software libre y de código abierto.

El sistema de diálogo desarrollado para el proyecto DIHANA se evaluó a partir de la interacción con 225 usuarios, cada uno de los cuales llevó a cabo un total de 4 diálogos. Cada uno de ellos debía responder un cuestionario en el que indicaba su opinión personal al utilizar el sistema. La Figura 3 muestra las estadísticas de las preguntas respondidas por los usuarios.

¿Entendió al sistema cuando éste hablaba?				
Siempre	Habitualmente	La mayor parte de veces	A veces	Nunca
72,51%	20,85%	6,64%	0%	0%
¿El sistema comprendió lo que usted le decía?				
Siempre	Habitualmente	La mayor parte de veces	A veces	Nunca
4,74%	57,82%	33,18%	4,27%	0%
¿Fue adecuado el ritmo de la interacción?				
Siempre	Habitualmente	La mayor parte de veces	A veces	Nunca
9%	44,55%	28,44%	15,17%	2,84%
¿Supo usted cómo actuar en cada momento del diálogo?				
Siempre	Habitualmente	La mayor parte de veces	A veces	Nunca
38,86%	35,07%	24,64%	1,42%	0%
¿Con qué frecuencia el sistema fue lento en su respuesta?				
Siempre	Habitualmente	La mayor parte de veces	A veces	Nunca
6,16%	12,80%	23,70%	55,45%	1,9%
¿El sistema se comportó del modo esperado durante la conversación?				
Siempre	Habitualmente	La mayor parte de veces	A veces	Nunca
14.69%	50.24%	31.75%	3.32%	0%

**Figura 3.** Estadísticas de las preguntas formuladas a los usuarios que utilizaron el sistema DIHANA

## 4.2. El sistema MYVOICE

Los sistemas de diálogo MyVoice y MyDictate (Cerva y Nouza, 2007) fueron desarrollados en la Technical University of Liberec para posibilitar el acceso a diferentes aplicaciones mediante la voz. MyVoice permite que personas con discapacidad motora puedan trabajar con un ordenador empleando varios centenares de comandos orales. Para cumplir este propósito, MyVoice interpreta dichos comandos como una o varias acciones básicas a llevar a cabo; por ejemplo, presionar, mantener o soltar una tecla o combinación de teclas, mover el puntero del ratón y presionar los botones del mismo, ejecutar programas, e imprimir secuencias de caracteres. MyDictate es el primer programa de dictado desarrollado para el idioma checo y dispone de un gran vocabulario que contiene las 540.000 palabras checas más frecuentes.

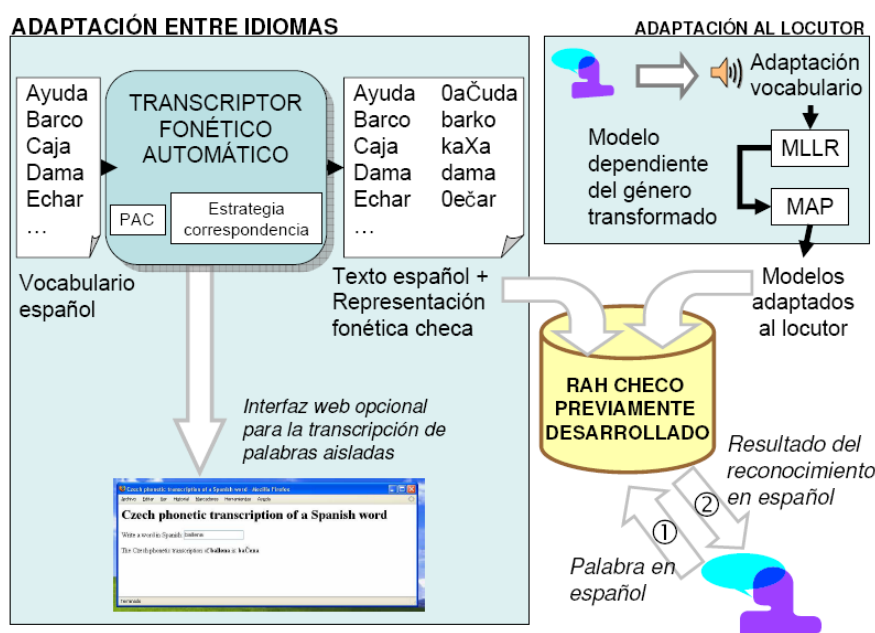
El reconocedor del habla checo incluido en el sistema es fruto del trabajo desarrollado durante más de una década en la Technical University of Liberec (Nouza et al., 2005). Sus modelos acústicos se basan en modelos ocultos de Markov con unidades del habla independientes del contexto que son proporcionadas por varios modelos de ruido, con distribuciones de salida que utilizan al menos 64 gaussianas. Dependiendo de las



condiciones de aplicación, estos modelos pueden ser independientes del locutor (speaker independent, SI), adaptados al género (gender dependent, GD), o adaptados al locutor (speaker adapted, SA). Por otra parte, el módulo de decodificación integrado en el reconocedor utiliza un léxico de palabras ordenadas alfabéticamente, cada una de las cuales se representa mediante su forma escrita y fonética.

El software MyVoice es utilizado en la actualidad por 60 usuarios discapacitados de la República Checa, cuyos informes demuestran que la tasa de palabras erróneas (word error rate, WER) oscila entre el 1% y 2 %, para los casos en los que el locutor no posee discapacidades en el habla. Uno de los estudios más importantes desarrollado con este sistema consistió en llevar a cabo la adaptación de su reconocedor para permitir la interacción con el mismo utilizando el español (Callejas et al., 2007).

MyVoice se estructura en varios grupos de comandos, cada uno de los cuales se encarga de una tarea específica, por ejemplo el grupo que controla el ratón es diferente del que gestiona el teclado, pero puede accederse fácilmente a todos los demás desde cada uno de ellos usando comandos de voz. El tamaño de estos grupos varía entre 5 y 137 comandos, donde el más grande contiene principalmente los nombres de las letras del alfabeto, y de las teclas del PC, lo que hace que el reconocimiento sea muy difícil debido a que la diferencia acústica entre ellas es muy sutil. Sin embargo, dado que se ha definido un vocabulario específico para cada tarea, se consiguen mejores resultados de reconocimiento cuando se agrupan los comandos. Además, este agrupamiento facilita la interacción, puesto que el usuario es consciente de las palabras válidas que pueden pronunciarse en cualquier momento. El proceso de adaptación entre idiomas llevado a cabo se muestra en la Figura 4. Los resultados obtenidos tras aplicar una técnica de correspondencia fonética muestran que es posible obtener WER menores al 6% en la utilización on-line en castellano con My Voice mediante los comandos definidos.

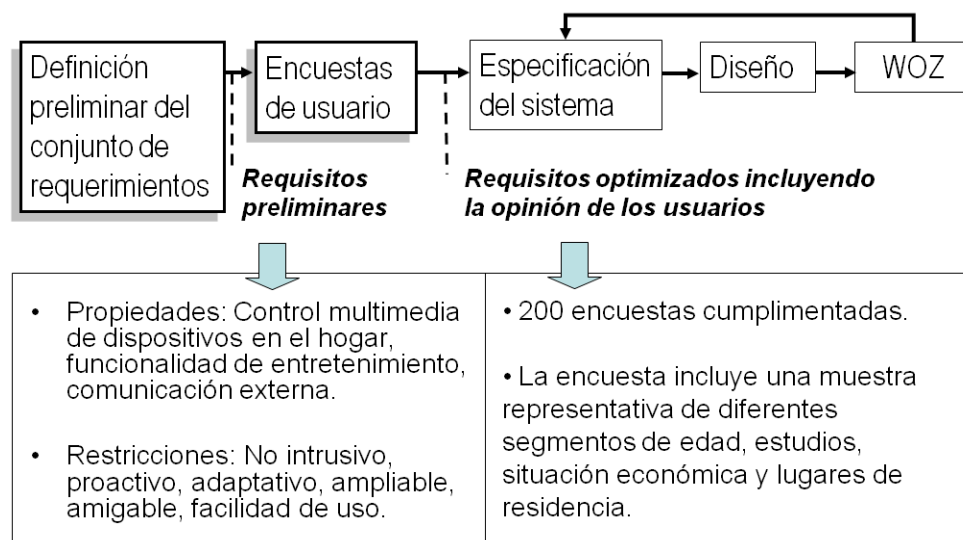


**Figura 4.** Proceso propuesto para la adaptación entre idiomas

### 4.3. Sistemas de diálogo e inclusión social

En último lugar se presenta un estudio para determinar las funcionalidades a incorporar en el diseño de un sistema de diálogo para la asistencia a personas mayores en el hogar (Callejas et al., 2006). Para ello, se diseñaron un conjunto de cuestionarios en los que se pregunta a los usuarios potenciales de la aplicación sobre qué funcionalidades incorporaría en el sistema, qué modalidades de interacción incluiría y qué aspectos valoraría más en los diferentes servicios llevados a cabo por el sistema. El cuestionario fue respondido por 200 hombres y mujeres con edades comprendidas entre 50-80 de diferentes poblaciones de la provincia de Granada, incluyéndose personas con diferentes discapacidades. La Figura 5 muestra el proceso diseñado

para llevar a cabo la implementación del sistema final teniendo en cuenta la opinión mostrada por los usuarios a través de estas encuestas.



**Figura 5.** Desarrollo final de un sistema de diálogo teniendo en cuenta la opinión previa de los usuarios.

A partir de los resultados obtenidos en esta encuestas se determinó la importancia de centrar los esfuerzos de diseño del sistema en determinadas funcionalidades (control de las persianas y ventanas, dispositivos para la comunicación con otras personas, etc.), así como la importancia de diseñar dichas funcionalidades teniendo en cuenta las preferencias mostradas por los diferentes colectivos

## 5. Conclusiones

El diseño de sistemas de diálogo orales es uno de los campos más atractivos y en el que se desarrolla actualmente una mayor investigación dentro de los campos del procesamiento del lenguaje natural y las tecnologías del habla. Desarrollar interfaces orales inteligentes posibilita dos objetivos fundamentales. En primer lugar, permiten el acceso a aplicaciones en entornos en los que el uso de los interfaces tradicionales imposibilitaría su utilización (por ejemplo, acceso a aplicaciones desde el automóvil). En segundo lugar, posibilitan el acceso a dichas aplicaciones a personas con discapacidades, haciendo uso de la voz para posibilitar su utilización.

A lo largo de esta comunicación se han presentado las características fundamentales de estos sistemas, su evolución y módulos principales que los componen. Con las aplicaciones mostradas se ha incidido en la utilización de estos sistemas para el desarrollo de aplicaciones adaptadas al usuario, el empleo de software libre para realizar su implementación y la importancia que poseen para facilitar el acceso a las aplicaciones a personas con discapacidades.

## Agradecimientos

Trabajo desarrollado en el marco del proyecto 08-117 “Aplicación de computación ubicua e interfaces multimodales para la convergencia al aprendizaje autónomo y adquisición de competencias en las tutorías”, Unidad de Innovación Docente, Universidad de Granada (España).

## REFERENCIAS BIBLIOGRÁFICAS

- Alexandersson, J., Becker, T. Overlay as the basic operation for discourse processing in a multimodal dialogue system. En: *Proc. of the 2<sup>nd</sup> IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*. Seattle, USA. 2001. P. 8-14.
- Andeani, G., Fabbizio, D. D., Gilbert, M., Gillick, D., Hakkani-Tur, D., Lemon, O. Let's Discoh: Collecting an Annotated Open Corpus with Dialogue Acts and Reward Signals for Natural Language Helpdesks. En: *Proc. of IEEE 2006 Workshop on Spoken Language Technology (SLT'06)*. Palm Beach, Aruba. 2006. P. 218-221.
- ARPA. Speech and Natural Language Workshop. En: *Book of Proceedings*, San Mateo (Estados Unidos). 1994.
- Aust, H., M. Oerder, F. Seide, V. Steinbiss. The Philips automatic train timetable information system. En: *Speech Communication*, 17. 1995. P. 249-262.
- Aust, H., M. Oerder, M. Seide, V. Steinbiss. The Philips automatic train timetable information system. En: *Proc. of the Interactive Voice Technology for Telecommunications Applications (IVTTA'94)*, Kyoto (Japón). 1994. P. 67-72.
- Baggia, P., G. Castagneri, M. Danieli. Field trials of the Italian ARISE train timetable system. En: *Speech Communication*, volumen 31, 2000. P. 355-367.
- Bernsen, N.O., L. Dybkjaer. The DISC project. En: *ELRA Newsletter*, volumen 2(2), 1997. P. 6-8.
- Billi, R., L.F. Lamel. Railtel: Railway telephone services. En: *Speech Communication*, volumen 23, 1997. P. 63-82.
- Bohus, D., Grau, S., Huggins-Daines, D., Keri, V., Krishna, G., Kumar, R., Raux, A., Tomko, S. Conquest - an Open-Source Dialog System for Conferences. En: *Proc. of Human Language Technologies'07: The Conference of the North American Chapter of the Association for Computational Linguistics*. Rochester, NY, USA. 2007. P. 9-12.
- Bohus, D., Rudnicky, A. Larri: A Language-Based Maintenance and Repair Assistant. En: *Proc. of Multi-Modal Dialogue in Mobile Environments Conference (IDS'02)*. Kloster Irsee, Germany. 2002. P. 203-218.
- Bos, Johan, Ewan Klein, Oliver Lemon, Tetsushi Oka. The verbmobil prototype system - a software engineering perspective. En: *Journal of Natural Language Engineering*, volumen 5(1), 1999. P. 95-112.
- Boves, L., Os, E. D. Multimodal services, a MUST for UMTS. Informe técnico, EURESCOM. 2002.
- Callejas, Z., Nouza, J., Cerva, P., López-Cózar, R. Myvoice goes Spanish. Cross-lingual adaptation of a voice-controlled PC tool for handicapped people. En: *Procesamiento del Lenguaje Natural* 39, 2007. P. 277-278.
- Callejas, Z., López-Cózar, R. Human-centred Development of Interactive Systems: Improving Usability in Early Lifecycle Stages. En: *Proc. of Interspeech-06 Satellite Workshop Dialogue on Dialogues. Multidisciplinary Evaluation of Advanced Speech-based Interactive Systems*, Pittsburgh (Estados Unidos). 2006.

- Castagneri, G., P.Baggia, M. Danieli. Field trials of the Italian ARISE train timetable system. En: *Proc. of the Interactive Voice Technology for Telecommunications Applications Workshop (IVTTA'98)*. 1998. P. 97-102.
- Catizone, R., Setzer, A., Wilks, Y. Multimodal Dialogue Management in the COMIC Project. En: *Proc. of EACL'03 Workshop on Dialogue Systems: interaction, adaptation, and styles of management*. Budapest, Hungary, 2003. P. 25-34.
- Cerva, P., Nouza, J. Design and development of voice controlled aids for motor-handicapped persons. En: *Proc. of the 11th International Conference on Spoken Language Processing (Interspeech'07-Eurospeech)*. Antwerp, Belgium, 2007. P. 2521-2524.
- Cole, R., Hirschman, L. [et.al]. The Challenge of Spoken Language Systems: Research Direction for the Nineties. En: *IEEE Transactions on Speech and Audio Processing* 3, 1995. P. 1-20.
- Corradini, A., [et.al]. Towards believable behavior generation for embodied conversational agents. En: *Proc. of the Workshop on Interactive Visualisation and Interaction Technologies (IV&IT 2004)*. Krakow, Poland, 2004. P.946-953.
- Dale, R. Next-generation spoken dialog systems. *Technology Trends Seminar Series*, Macquarie University, 2003. [Consulta: 9/09/2009]. Disponible desde Internet: <http://www.ict.csiro.au/MU/Trends/2003.htm>
- Darpa. Speech and Natural Language Workshop. En: *Book of Proceedings*, San Mateo (Estados Unidos), 1992.
- Gao, Y., Gu, L., Jeff, H.K. Portability challenges in developing interactive dialogue systems. En: *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*. Vol. 5. Philadelphia, PA, USA, 2005. P.18-23.
- Gauvain, J.L., S. Bennacef, L. Devillers, y L. Lamel. Spoken language system development for the Mask Kiosk. En: *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU'95)*, Salt Lake City (Estados Unidos). 1995. P.119-120.
- Griol, D., Hurtado, L. F., Segarra, E., Sanchis, E. A statistical approach to spoken dialog systems design and evaluation. *Speech Communication*, 50(8-9). 2008. P. 666-682.
- Haas, J., Gallwitz, F., Horndasch, A., Huber, R., Warnke, V., 2005. Telephone-based speech dialog systems. *Lecture Notes in Computer Science* 3663, P. 125-132.
- Hardy, H., Biermann. [et.al]. The Amitiés system: Data-driven techniques for automated dialogue. *Speech Communication* 48, 2006. P. 354-373.
- Hubal, R. C., Frank, G. A., Guinn, C. I.,. Avatalk Virtual Humans for Training with Computer Generated Forces. En: *Proc. of the 9th Conference on Computer Generated Forces and Behavioral Representation*. Orlando, Florida, USA, 2000. P. 617-623.
- Jokinen, K.,. Natural interaction in spoken dialogue systems. En: *Proc. of the Workshop Ontologies and Multilinguality in User Interfaces*. Crete, Greece, 2003. P.730-734.
- Jokinen, K., Kanto, K., Rissanen, J. Adaptive User Modelling in AthosMail. *Lecture Notes in Computer Science* 3196, 2004. P. 149-158.

- Kacic, Z. Advances in spoken dialogue systems development. En: *Proc. of IEEE International Symposium on Industrial Electronics (ISIE'99)*. Vol. 1. Bled, Slovenia. 1999. P.169-172.
- Krsmanovic, F., Spencer, C., Jurafsky, D., Ng, A. Y. Have we meet? MDP Based Speaker ID for Robot Dialogue. En: *Proc. of the 9th International Conference on Spoken Language Processing (Interspeech'06-ICSLP)*. Pittsburgh, USA, 2006. P. 461-464.
- Lamel, L. [et.al]. The LIMSI ARISE system. *Speech Communication*, volumen 31, 2000. P. 339-353.
- Lemon, O., Bracy, A., Gruenstein, A., Peters, S. The Witas Multi- Modal Dialogue System. En: *Proc. of Eurospeech'01*. Aalborg, Denmark, 2001. P.1559-1562.
- Lemon, O., Georgila, K., Henderson, J. Evaluating Effectiveness and Portability of Reinforcement Learned Dialogue Strategies with real users: the TALK TownInfo Evaluation. En: *Proc. of IEEE-ACL Workshop on Spoken Language Technology (SLT 2006)*. Palm Beach, Aruba, 2006. P. 178-181.
- Litman, D. J., Silliman, S. Itspoke: An Intelligent Tutoring Spoken Dialogue System. En: *Proc. of the Human Language Technology Conference: 4th Meeting of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*. Boston, USA, 2004. P. 233-236.
- Mattasoni, M. [et.al]. On the Joint Use of Noise Reduction and MLLR Adaptation for In-Car Hands-Free Speech Recognition. En: *Proc. of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'02)*. Orlando, USA, 2002. P. 289-292.
- Melin, H., Sandell, A., Ihse, M. CTT-bank: A speech controlled telephone banking system - an initial evaluation. *TMH-QPSR* 1, 2001. P. 1-27.
- Menezes, P., Lerasle, F., Dias, J., Germa, T. Towards an interactive humanoid companion with visual tracking modalities. En: Hackel, M. (Ed.), *Humanoid robots: Human-like machines*. Itech, 2007. P. 367-398.
- Minker, W. [et.al]. The Seneca Spoken Language Dialogue System. *Speech Communication* 43, 2004.
- Minker, W. [et.al]. Recent trends in spoken language dialogue systems. En: *Proc. of ITI 4th International Conference on Information and Communications (ICICT'06)*. Montreal, Canada, 2006a. P.1-16, invited paper.
- Minker, W. [et.al]. Next-generation human-computer interfaces - Towards intelligent, adaptive and proactive spoken language dialogue systems. En: *Proc. of the 2nd IET International Conference on Intelligent Environments (IE'06)*. Vol. 1. Athens, Greece, 2006B. P. 213-219.
- Montoro, G., Alamán, X., Haya, P.A. Spoken interaction in intelligent environments: a working system. En: Ferscha, A., Hoertner, H., Kotsis, G. (Eds.), *Advances in Pervasive Computing*. Austrian Computer Society (OCG), 2004. P. 747-754.
- Montoro, G. [et.al]. A proposal for an XML definition of a dynamic spoken interface for ambient intelligence. En: *Proc. of International Conference on Intelligent Computing (ICIC'06)*. Kunming, China, 2006. P. 711-716.
- Mostow, J., 2008. Experience from a reading tutor that listens: Evaluation purposes, excuses, and methods. En: Kinzer, C., Verhoeven, L. (Eds.), *Interactive Literacy Education: Facilitating Literacy Environments Through Technology*. New York: Lawrence Erlbaum Associates, Taylor and Francis, P. 117-148.
- Nouza, J., T., Cerva, P. A multi-functional voice-control aid for disabled persons. En: *Proc. of International Conference on Speech and Computer (SPECOM'05)*. Patras, Greece, 2005. P. 715-718.

- Peckham, J. A new generation of spoken dialogue systems: results and lessons from the SUNDIAL project. En: *Proc. of European Conference on Speech Communications and Technology (Eurospeech'93)*, volumen 1, Berlín (Alemania). 1993 P. 33-42,
- Pérez, G., Amores, G., Manchón, P. A multimodal architecture for home control by disabled users. En: *Proc. of IEEE/ACL Workshop on Spoken Language Technology (SLT)*. Palm Beach, Aruba. 2006. P. 134-137.
- Pieraccini, R., E. Levin, Y W. Eckert. Amica: The AT&T mixed initiative conversational architecture. En *Proc. of European Conference on Speech Communications and Technology (Eurospeech'97)*, Rodas (Grecia). 1997. P.1875-1878
- Rayner, M. [et.al]. A voice enabled procedure browser for the International Space Station. En: *Proc. of 43th Annual Meeting of the Association for Computational Linguistics*. Ann Arbor, USA, 2005. P. 29-32.
- Raux, A., Langner, B., Black, A., Eskenazi, M. Let's Go Public! Taking a Spoken Dialog System to the Real World. En: *Proc. of the 9th International Conference on Spoken Language Processing (Interspeech'05-Eurospeech)*. Lisbon, Portugal, 2005. P. 885-888.
- Rich, C., Sidner, C. Collagen: A Collaboration Manager for Software Interface Agents. *User Modeling and User-Adapted Interaction* 8, 1998. P. 315-350.
- Roque, a. . [et.al]. 2006. Radiobot-CFF: A Spoken Dialogue System for Military Training. En: *Proc. of the 9th International Conference on Spoken Language Processing (Interspeech'06-ICSLP)*. Pittsburgh, USA, P.477-480.
- Rudnicky, A. [et.al]. Creating natural dialogs in the Carnegie Mellon Communicator system. En: *Proc. of European Conference on Speech Communications and Technology (Eurospeech'99)*, volumen 1(4), 1999. P. 1531-1534.
- Seneff, S., Polifroni, J., 2000. Dialogue management in the Mercury flight reservation system. En: *Proc. of ANLP-NAACL Workshop on Conversational systems*. Vol. 3. Seattle, Washington, USA, P.11-16.
- Seneff, S., L. Hirschman, Y V. Zue. 1991. Interactive problem solving and dialogue in the ATIS domain. En *Proc. of the Fourth ARPA Speech and Natural Language Workshop*, P.354-359, Pacific Grove, California (Estados Unidos).
- Skantze, G., Edlund, J., Carlson, R. Talking with Higgins: Research challenges in a spoken dialogue system. En: *Proc. of Perception and Interactive Technologies (PIT'06)*. Kloster Irsee, Germany, 2006. P.193-196.
- Stewart, J. An electrical analog of the vocal tract. *Nature* 110, 1922. P. 311-312.
- Turing, A. Computing machinery and intelligence. *Mind* 236, 1950. P. 433-460.
- Vaquero, C. [et.al]. An application for computer-aided speech therapy in spanish language. En: *Proc. of IV Jornadas en Tecnología del Habla*. Zaragoza, Spain, 2006. P. 321-326.
- Villing, J., Larsson, S. Dico - A Multimodal Menu-based In-vehicle Dialogue System. En: *Proc. of the 10th Workshop on the Semantics and Pragmatics of Dialogue (BRANDIAL'06)*. Potsdam, Germany, 2006. P.187-188.

- Ward, W. Y B. Pellom. The CU Communicator system. En: *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU'99)*, Keystone, Colorado (Estados Unidos), 1999. P. 341-344.
- Weizenbaum, J. Eliza - A computer program for the study of natural language communication between man and machine. En: *Communications of the ACM* 9 (1), 1966. P. 36-45.
- Weng, F. [et.al]. CHAT: A Conversational Helper for Automotive Tasks. En: *Proc. of the 9th International Conference on Spoken Language Processing (Interspeech'06-ICSLP)*. Pittsburgh, USA, 2006. P.1061-1064.
- Zue, V. [et.al]. The MIT ATIS System: December 1993 Progress Report. En: *Proc. of ARPA Spoken Language Technology Workshop*, Princeton (Estados Unidos). 1993. P. 339-353.
- Zue, V. W., Glass, J. R.,. Conversational Interfaces: Advances and Challenges. En: *Proc. of the IEEE* 88 (8), 2000. P. 1166-1180.